AD_____

Award Number: DAMD17-99-1-9119

TITLE: A New Model for the Estimation of Breast Cancer Risk

PRINCIPAL INVESTIGATOR: Maryellen L. Giger, Ph.D.

CONTRACTING ORGANIZATION: The University of Chicago
                          Chicago, Illinois 60637

REPORT DATE: July 2001

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
              Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
                        Distribution Unlimited

The views, opinions and/or findings contained in this report are
those of the author(s) and should not be construed as an official
Department of the Army position, policy or decision unless so
designated by other documentation.

20020124 313

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE July 2001	3. REPORT TYPE AND DATES COVERED Annual (14 Jul 00 - 13 Jul 01)	
4. TITLE AND SUBTITLE A New Model for the Estimation of Breast Cancer Risk		**5. FUNDING NUMBERS** DAMD17-99-1-9119	
6. AUTHOR(S) Maryellen L. Giger, Ph.D.			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) The University of Chicago Chicago, Illinois 60637 E-Mail: m-giger@uchicago.edu		**8. PERFORMING ORGANIZATION REPORT NUMBER**	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012		**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**	

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited	12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 Words)*

Cancer risk is the probability that cancer will occur in a given population. Research on cancer risk seeks to identify populations with a high probability of developing cancer. The goal of this research is to merge a computerized analysis of mammograms, which characterizes the breast pattern, with information of a woman's personal and family histories into a novel model for use in estimating risk of breast cancer. The specific aims include 1. Creating a database of mammograms, along with tabulated clinical information of women at low risk and high risk for breast cancer; 2. Developing a new model using computer methods for merging mammographic information with clinical information; and 3. Evaluating the efficacies of the new model compared to currently used methods of risk assessment. The main hypothesis to be tested is that given a group of women, the new computerized risk model that merges computerized analyses of mammograms with clinical information should yield a novel way for identifying those women at risk for breast cancer.

To date, we have shown that computer-extracted features of mammographic parenchymal patterns can be used in the prediction of breast cancer risk. This has been demonstrated (on the developing database) using three approaches: (1) correlation with clinical models of Gail and Claus, (2) separation between women at low risk and those with a positive gene testing result, and (3) separation between women at low risk and those that have breast cancer. In addition, we have shown, in a preliminary study, that the inclusion of the mammographic features with age increase the predictive power over the use of age alone in the prediction of breast cancer risk.

14. SUBJECT TERMS breast cancer			15. NUMBER OF PAGES 17
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** Unlimited

NSN 7540-01-280-5500

Table of Contents

Proposal Title: A New Model for the Estimation of Breast Cancer Risk

P.I.: Maryellen L. Giger, Ph.D.

INTRODUCTION:

Cancer risk is the probability that cancer will occur in a given population. Research on cancer risk seeks to identify populations with a high probability of developing cancer. The goal of this research is to merge a computerized analysis of mammograms, which characterizes the breast pattern, with information of a woman's personal and family histories into a novel model for use in estimating risk of breast cancer.

The specific aims include 1. Creating a database of mammograms, along with tabulated clinical information of women at low risk and high risk for breast cancer; 2. Developing a new model using computer methods for merging mammographic information with clinical information; and 3. Evaluating the efficacies of the new model compared to currently used methods of risk assessment. The main hypothesis to be tested is that given a group of women, the new computerized risk model that merges computerized analyses of mammograms with clinical information should yield a novel way for identifying those women at risk for breast cancer. It should be noted that current clinical methods of assessing risk using the Gail or Claus models (clinical data only) are limited as illustrated by our preliminary studies, which show only moderate correlation between these two current models for cumulative risk and 10-year risk.

The new model will include computer-extracted features from digitized mammograms and clinical information from each woman. The computer-extracted features will be extracted within regions of digitized mammograms. In general, the breast can be described by the amount of dense regions (a percent dense) and by the heterogeneity/homogeneity of the dense portion pattern (texture). In addition, clinical information such as age and reproductive history contribute to the determination of risk. Therefore, methods of combining clinical data and multiple mammographic markers into a single model of risk will be developed for the model.

Potential uses of this innovative model include 1) serving as a means to assess the cancer risk of women undergoing routine screening mammography and thus, identifying those women that may require closer scrutiny and 2) serving as a means to monitor the cancer risk of women undergoing chemoprevention treatments. The research is novel in that currently there does not exist a reliable means to assess the cancer risk of individual women

using both mammographic and clinical information. In addition, if a woman knew that she was at an increased risk of breast cancer, it is likely that she would better comply with screening mammography programs. In the future, a successful model could also be used to assess the effect of chemoprevention on a women's parenchymal pattern and thereby, overall risk.

BODY:

Task 1. Establishment of database (mos. 1-30)

The high-risk database is being collected within the University of Chicago Cancer Risk Clinic and consists of mammograms, pedigree information, epidemiological data and related biological specimens from patients with a family history of breast cancer. All mammograms done since 1990 are being collected for all participants irrespective of their cancer status. Breast Cancer risk assessment is performed using both Gail and Claus models and genetic testing whenever possible. A low-risk database is also being collected from our breast cancer screening program and includes mammograms and clinical information on women undergoing routine screening mammograms. The low risk database is being developed to include women who are age-matched to reflect the age of women in our high risk database. We have collected cases from over 100 patients and Gail and Claus calculations have been performed. We now have approximately 35 patients with positive BRCA1/BRCA2 gene mutation testing.

The mammograms are converted to digital format by using a laser film scanner (2048 by 2048 matrix with 12-bit quantization). Such high spatial resolution is necessary in order to adequately retain the high-frequency texture patterns.

Task 2. Development of risk model including mammographic markers and clinical information (mos. 3-30)

Computerized analysis of the parenchymal pattern is based on various texture analysis methods we have developed in our laboratory including Fourier spectra analysis, histogram analysis, and artificial neural networks. Fourteen features are currently extracted within the regions of each digitized mammogram. These features are grouped into (i) features based on the absolute values of the gray levels, (ii) features based on

5

gray-level histogram analysis, (iii) features based on the Fourier transform, and (iv) features based on the spatial relationship among gray levels.

The purpose of one of our studies, was to identify computer-extracted, mammographic parenchymal patterns that are associated with breast cancer risk. We extracted fourteen features from the central breast region on digitized mammograms to characterize the mammographic parenchymal patterns of women at different risk levels. In the study, the features were used to characterize mammographic patterns seen in low-risk women and in women who have breast cancer. Stepwise linear logistic regression was employed to identify useful features to differentiate between the mammographic patterns of low-risk women and women with breast cancer. The relationship between these mammographic patterns and the risk of developing breast cancer was identified based on the odds ratios associated with these individual features. We also employed two different approaches to relate these mammographic features to breast cancer risk. In one approach, the features were used to distinguish mammographic patterns seen in low-risk women from those who inherited a mutated form of the *BRCA1/BRCA2* gene. In another approach, the features were related to risk as determined from existing clinical models (*Gail* and *Claus* models). Stepwise linear discriminant analysis was employed to identify features that were useful in differentiating between "low-risk" women and *BRCA1/BRCA2*-mutation carriers. Stepwise linear regression analysis was employed to identify useful features in predicting the risk as estimated from the *Gail* and *Claus* models. The computer-extracted mammographic features identified from this approach were similar to those identified from the two previous approaches. The results from this study show that women who have dense breasts and whose mammographic patterns are coarse and low in contrast have an increased risk of developing breast cancer. The consensus of the findings from the three different approaches substantiated the existing results. (Presented CARS 2000) Our features were further validated this year when we extended our number of cases from gene-carriers to 30. This resulted in a RSNA 2000 presentation (November, 2000) and a recently submitted paper to Radiology (accepted pending revision, July 2001).

We also analyzed the contributions of age and computer-extracted mammographic features in the prediction of breast cancer risk. We assessed the contribution of the computer-extracted features to risk prediction in terms of percent increase in the prediction power (r^2) when age (the single most important risk factor for breast cancer) was used alone and when the mammographic features were included. The inclusion of the mammographic features increased the prediction power (r2) from 0.08 and 0.16 (age alone) to 0.17 and 0.32, yielding an increase of 113% an d 100% in r2 for

6

predicting the risk as estimated from the Gail and Claus models. The substantial increase in r^2 indicates the important contribution of these mammographic features in risk prediction and the need to incorporate in predicting breast cancer risk. (Presented IWDM 2000)

Task 3. Evaluation methods

This task is planned for months 20-36. However, our plans include the following.

We are developing a model for assessing breast structure and cancer risk. Thus, correlation analysis will be used in evaluating the performance of the measures. Linear correlation analysis will be performed to determine the correlation among the output of the new model and the Gail risk model (or Claus model). We are using the combined model based on the first two models (gene mutation vs. low-risk and with cancer vs. without cancer) and evaluating the performance of the combined measures using the Gail model.

Another task for the coming year will be to evaluate the texture measures in their ability to predict the onset of breast cancer (over time). Based on the cases collected during the first 2.5 years of the study, a nested case-control study design will be implemented. As our criteria are that the mammograms should have been obtained after 1989, there is potential for collecting images from eight years ago (so can assume 5 to 8 year follow-up). In a nested case-control database, the cases will correspond to women who will have developed cancer and the control will correspond to women who will have stayed cancer free during the period. We will calculate the clinical markers (e.g., Gail) and the mammographic features of the initial examination prior to the 5 to 8 year follow-up. Multivariate analysis will be used to examine the relationship between the new model and risk of breast cancer while controlling for other risk factors such as age at menarche and parity. A proportional-hazards regression model will be used to calculate the relative risk for each radiographic marker.

KEY RESEARCH ACCOMPLISHMENTS:

- Further increase in our database of high and low risk cases, especially those with positive BRCA1/BRCA2 testing.
- Further verification of our texture features for characterizing the breast parenchyma using three different approaches -- all yielding the same result

7

- Preliminary study looking at the contribution of age and mammographic features to breast cancer risk prediction.

REPORTABLE OUTCOMES:

1.   Analysis of the relative contributions of mammographic features and age to breast cancer risk prediction. Zhimin Huo, Maryellen L. Giger and Olufunmilayo I. Olopade, **Presentation** at International Workshop on Digital Mammography 2000 (Toronto, Canada)

2.   Computerized analysis of mammographic patterns of women with and without breast cancer. Zhimin Huo, Maryellen L. Giger and Olufunmilayo I. Olopade, **Presentation** at CARS 2000 (San Fransico, CA)

3.   Huo Z, Giger ML, Wolverton DE, Zhong W, Cummings S, Olopade OI: Computerized analysis of mammographic parenchymal patterns for breast cancer risk assessment: Feature selection. **Journal article** Medical Physics 27:4-12, 2000.

4.   Huo Z, Giger ML, Zhong W, Nishikawa, RE, Wolverton DE, Olopade OI: "Mammographic parenchymal patterns as predictors for breast cancer risk". **Presentation** at 86th Scientific Assembly and Annual Meeting of Radiological Society of North America, Chicago, Illinois, 2000.

CONCLUSIONS:

To date, we have shown that computer-extracted features of mammographic parenchymal patterns can be used in the prediction of breast cancer risk. This has been demonstrated (on the developing database) using three approaches: (1) correlation with clinical models of Gail and Claus, (2) separation between women at low risk and those with a positive gene testing result, and (3) separation between women at low risk and those that have breast cancer. In addition, we have shown, in a preliminary study, that the inclusion of the mammographic features with age increase the predictive power over the use of age alone in the prediction of breast cancer risk.

Computerized analysis of mammographic parenchymal patterns for breast cancer risk assessment: Feature selection

Zhimin Huo,[a)] Maryellen L. Giger,[a)] Dulcy E. Wolverton, and Weiming Zhong
Kurt Rossmann Laboratories for Radiologic Image Research, Department of Radiology,
5841 South Maryland Avenue, The University of Chicago, Chicago, Illinois 60637

Shelly Cumming and Olufunmilayo I. Olopade
Department of Hematology and Oncology, The University of Chicago, Chicago, Illinois 60637

Our purpose in this study was to identify computer-extracted, mammographic parenchymal patterns that are associated with breast cancer risk. We extracted 14 features from the central breast region on digitized mammograms to characterize the mammographic parenchymal patterns of women at different risk levels. Two different approaches were employed to relate these mammographic features to breast cancer risk. In one approach, the features were used to distinguish mammographic patterns seen in low-risk women from those who inherited a mutated form of the BRCA1/BRCA2 gene, which confers a very high risk of developing breast cancer. In another approach, the features were related to risk as determined from existing clinical models (*Gail* and *Claus* models), which use well-known epidemiological factors such as a woman's age, her family history of breast cancer, reproductive history, etc. Stepwise linear discriminant analysis was employed to identify features that were useful in differentiating between "low-risk" women and BRCA1/BRCA2-mutation carriers. Stepwise linear regression analysis was employed to identify useful features in predicting the risk, as estimated from the *Gail* and *Claus* models. Similar computer-extracted mammographic features were identified in the two approaches. Results show that women at high risk tend to have dense breasts and their mammographic patterns tend to be coarse and low in contrast. © *2000 American Association of Physicists in Medicine.* [S0094-2405(00)01001-4]

Key words: breast cancer risk, gene mutation, mammographic parenchyma, computerized classification, linear discriminant analysis, linear regression analysis

I. INTRODUCTION

Breast cancer is the most frequently diagnosed malignancy after skin cancer among women in the United States.[1] It is estimated that approximately one in eight women will be diagnosed with breast cancer in her lifetime.[1] Studies show that screening mammography is the best imaging technique for the early detection of breast cancer,[2,3] which reduces breast cancer deaths by as much as 30%.[4–9] Annual screening mammography has been recommended by the American Cancer Society for all women over the age of 40.[1]

With the increasing awareness of breast cancer risk and the benefit of screening mammography, more women in all risk categories are seeking information regarding their individual risk of developing breast cancer. Identification and close surveillance of women who are at high risk of developing breast cancer may provide an opportunity for early cancer detection.

Large-scale epidemiological studies have shown that, in addition to age, there are many factors associated with breast cancer risk, although the basic mechanisms underlying the association between breast cancer and these risk factors are not well understood. These include risk factors such as a woman's family history of breast cancer, her reproductive history, and her history of previous breast biopsies. Clinical models, such as the *Gail et al.* model[10] and the *Claus et al.* model,[11] have been developed to estimate an individual's

risk of developing breast cancer using these factors. Estimates of risk from these models have been used by clinicians for counseling women who are seeking information regarding their individual breast cancer risk.[12,13]

Recent molecular studies demonstrate that breast cancer may be inherited.[14–16] Genes that are responsible for inherited breast cancer, including the BRCA1 (breast cancer 1) and BRCA2 (breast cancer 2) genes, have been identified.[17] Although hereditary breast cancers account for only 5%–10% of all breast cancers,[18,19] it is estimated that women who inherit a mutated form of the BRCA1 or BRCA2 gene have as much as a 56%–87% risk of developing breast cancer by age 70 years,[20,21] which is about 8 times higher than the lifetime risk for the general population. DNA tests for these genes offer a way to identify women who have hereditary breast cancer.

The association of breast cancer risk with mammographic parenchymal patterns has been investigated in the past. Increased mammographic density has been found to be associated with an increased risk of breast cancer. It has been shown in several studies that women with increased mammographic parenchymal density are at a four- to six-fold higher risk over women with primarily fatty breasts.[22–25] At present, the reason for this increased risk is unclear. One possibility is that increased density reflects a larger amount of tissue at risk for developing breast cancer. Since most

breast cancers develop from the epithelial cells that line the ducts of the breast, having more of this tissue as reflected by increased mammographic density may increase one's chances of developing breast cancer.

Wolfe first described a possible association between the risk for breast cancer and different mammographic patterns in 1976.[22] Since then, many investigators have used the Wolfe patterns to classify the mammographic appearance of breast parenchyma for risk assessment.[26] Others have used qualitative or quantitative estimates of the proportion of the breast area (percent dense) that mammographically appears dense to assess the associated breast cancer risk. Although considerable variations were observed in reported individual results based on visual assessment,[26] most studies showed that women with dense breasts have an increased risk of breast cancer relative to those with fatty breasts.

While visual assessment of mammographic patterns has remained controversial due to the subjective nature of human assessment,[27] computer vision methods can yield objective measures of breast density patterns. Computerized classification of mammographic images has been investigated by various investigators, including Magnin *et al.*,[28] Caldwell *et al.*,[29] and Tahoces *et al.*,[30] who used computer-extracted texture measures to classify mammographic patterns into the four categories of Wolfe patterns, and Taylor *et al.*[31] and Byng *et al.*,[32,33] who used computer-extracted texture features to quantify the percent dense of the breast. Byng *et al.*[32,33] first investigated the association of computer-extracted texture measures (i.e., skewness and fractal dimension) with breast cancer risk. They showed that increased mammographic density was associated with an increased relative risk of 2 to 4.

Our objective in this study is to identify computer-extracted mammographic features on digitized mammograms that are associated with breast cancer risk.[34] A total of 14 mammographic features from the central breast region were extracted. In general, breast parenchymal can be described by the amount of dense regions and by the heterogeneity/homogeneity of the patterns in the dense portions of the breast. We based our computer-extracted features on those that are already known to be associated with breast cancer risk from visual assessment.[22,24,25] Some of these individual computer-extracted features quantify percent dense while others characterize the heterogeneity. We believe that a combination of multiple features will perform better than a single feature in characterizing mammographic patterns, and thus may help in assessing breast cancer risk. These features were related to predictors of breast cancer risk using two different approaches: (1) the classification of mammographic patterns of low-risk women and BRCA1/BRCA2 gene-mutation carriers; and (2) the prediction of risk as estimated from the *Gail* model and the *Claus* model. The useful features were identified via the two different approaches. The characteristic mammographic patterns of women at high risk and at low risk were identified in terms of computer-extracted features.

II. MATERIALS AND METHODS

A. Database

Mammograms from 341 women were retrospectively collected. Information regarding women's reproductive histories, family histories of breast cancer, and histories of previous breast biopsies were collected to assess each individual's breast cancer risk using the *Gail* model and/or *Claus* model.[10,11] The information required by the *Gail* model in the calculation of individual risk are (1) age, (2) age at menarche, (3) age at first full-term birth, (4) number of first-degree relatives with breast cancer, and (5) number of previous breast biopsies. The information required by the *Claus* model in the calculation of individual risk are (1) age and (2) the number of first-degree and second-degree relatives with breast cancer and their ages of onset. Based on the calculated risk, 341 women were categorized into low-, moderate-, and high-risk groups. In addition, mammograms were collected from 15 women with BRCA1/BRCA2 mutation.

Mammograms from 285 of the women were obtained from the screening mammography program in the Department of Radiology (May 1996 to December 1996) at the University of Chicago Hospitals. These women completed questionnaires yielding information on their medical history and information required in the *Gail* or the *Claus* model. Their Gail and Claus risk estimates were calculated at the University of Chicago Cancer Risk Clinic (UCCRC), where genetic counseling is also performed. To be considered low risk in the study, women had to have no family history (no Claus risk) of breast cancer and the risk of developing breast cancer as estimated from the *Gail* model had to be less than 10%. Among the 285 women, 143 of them were considered to be low risk based on these criteria.

The 15 BRCA1/BRCA2 mutation carriers and an additional 56 women who were at high risk were recruited from the UCCRC. Mammograms previously obtained were retrieved and digitized for all these women. Information regarding their reproductive histories, family histories of breast cancer, histories of previous breast biopsies, etc. were collected to analyze their risk of developing breast cancer at the time of counseling. The mutation carriers were tested at a CLIA-approved laboratory under an IRB-approved protocol. Among the 15 BRCA1/BRCA2-mutation carriers, four had no cancer, two were diagnosed with ovarian cancer, and nine were diagnosed with breast cancer. For those with a previous diagnosis of breast cancer, mammograms obtained a year prior to the diagnosis were analyzed. These mammograms were reviewed by an expert mammographer and deemed void of any detectable abnormalities.

Since two analyses were performed in our study, the database (mammograms from the 356 women) were grouped as follows. Mammograms of the 143 low-risk women and the 15 mutation carriers were used in the *classification analysis.* Mammograms of the 341 women, excluding the 15 BRCA1/BRCA2-mutation carriers, were used in the *correlation analysis.* The BRCA1/BRCA2-mutation carriers were not included in the correlation study, because neither the *Gail* model nor the *Claus* model is accurate in predicting risk for
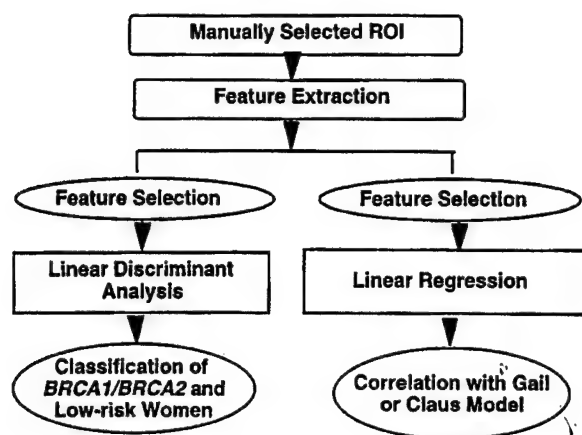
FIG. 1. The overall computerized scheme for breast cancer risk assessment.

FIG. 2. Digitized mammograms (cranial–caudal view) and a selected ROI.

women who are BRCA1/BRCA2 mutation carriers.[12]

It should be noted that the BRCA1/BRCA2-mutation carriers tend to be younger than the ''low-risk'' cases. The age of the BRCA1/BRCA2-mutation carriers ranged from 33 to 54 years, with a mean of 40.8 years and a median of 40 years. The age of the women in the low-risk group ranged from 35 to 54 years, with a mean of 44.7 years and a median of 45 years. To rule out possible bias due to the difference in age distribution of the BRCA1/BRCA2-mutation carriers and the ''low-risk'' women, classification was also performed on the 15 BRCA1/BRCA2 mutation carriers and 30 ''low-risk'' women who were randomly selected and age matched with the 15 BRCA1/BRCA2-mutation carriers at 5 year intervals. The two-to-one ratio of the number of low-risk women to that of the BRCA1/BRCA2-mutation carriers was determined, based on the number of age-matched cases available in the low-risk group.

B. Computerized analysis of parenchymal patterns on digitized mammograms

Figure 1 schematically outlines the computerized methods by which we investigated mammographic parenchymal patterns that are associated with breast cancer risk. Mammograms were digitized using a Konica laser scanner (LD 4500; Konica Medical, Wayne, NJ) at 0.1 mm pixel size and 10-bit gray-level scale. After digitization, regions-of-interest (ROIs), 256 pixels by 256 pixels in size, were manually selected from the central breast region (immediately behind the nipple). Figure 2 illustrates an example of a ROI selected from a digitized mammogram. The small ROI size (256 pixels by 256 pixels) was chosen in order to include small-sized breasts. ROIs selected from the central breast region behind the nipple were used for this study, because they usually include the most dense parts of the breast. It should be noted that in this study, a constant ROI size was used for all breast images regardless of breast size. ROIs were selected such that regions along the skin line that contains subcutaneous fat were not included.
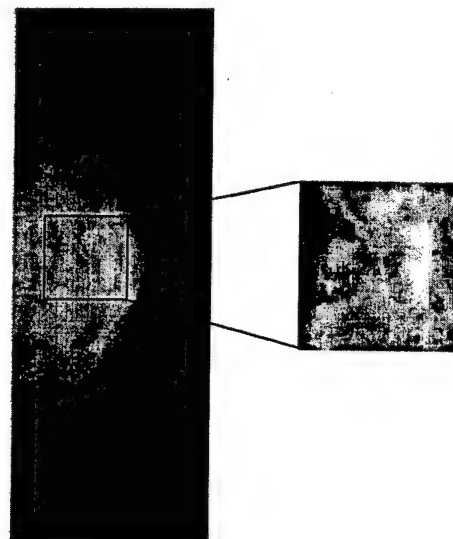
1. Computer-extracted features

A total of 14 features were extracted from each of the selected ROIs to quantitatively characterize the mammographic parenchymal patterns. These features are grouped into (i) features based on the absolute values of the gray levels, (ii) features based on gray-level histogram analysis, (iii) features based on the spatial relationship among gray levels within the ROI, and (iv) features based on Fourier analysis.

a. Features based on the absolute value of the gray levels. Features based on the absolute gray level values (features 1–7 below) include the maximum, the minimum, the average gray level, and various gray-level thresholds that yield 5%, 30%, 70%, and 95% of the area under the gray-level histogram of a ROI, as shown in Fig. 3. Figure 3 shows gray-level histograms of (a) a dense ROI, (b) a mixed ROI, and (c) a fatty ROI. Radiographically, the breast consists primarily of two types of tissue: fibroglandular tissue and fat. Regions of brightness in mammography associated with fibroglandular tissue are referred to as mammographic density. Features 1–7 are used as a means to quantify indirectly the brightness of the selected region, thus yielding information regarding the denseness of the region.

 (1) *MAX*: Maximum gray level of the ROI.
 (2) *MIN*: Minimum gray level of the ROI.
 (3) *AVG*: Average gray level of the ROI.
 (4) *5% threshold*: Gray level yielding 5% of the area under the histogram of the ROI.
 (5) *30% threshold*: Gray level yielding 30% of the area under the histogram of the ROI.
 (6) *70% threshold*: Gray level yielding 70% of the area under the histogram of the ROI.
 (7) *95% threshold*: Gray level yielding 95% of the area under the histogram of the ROI.

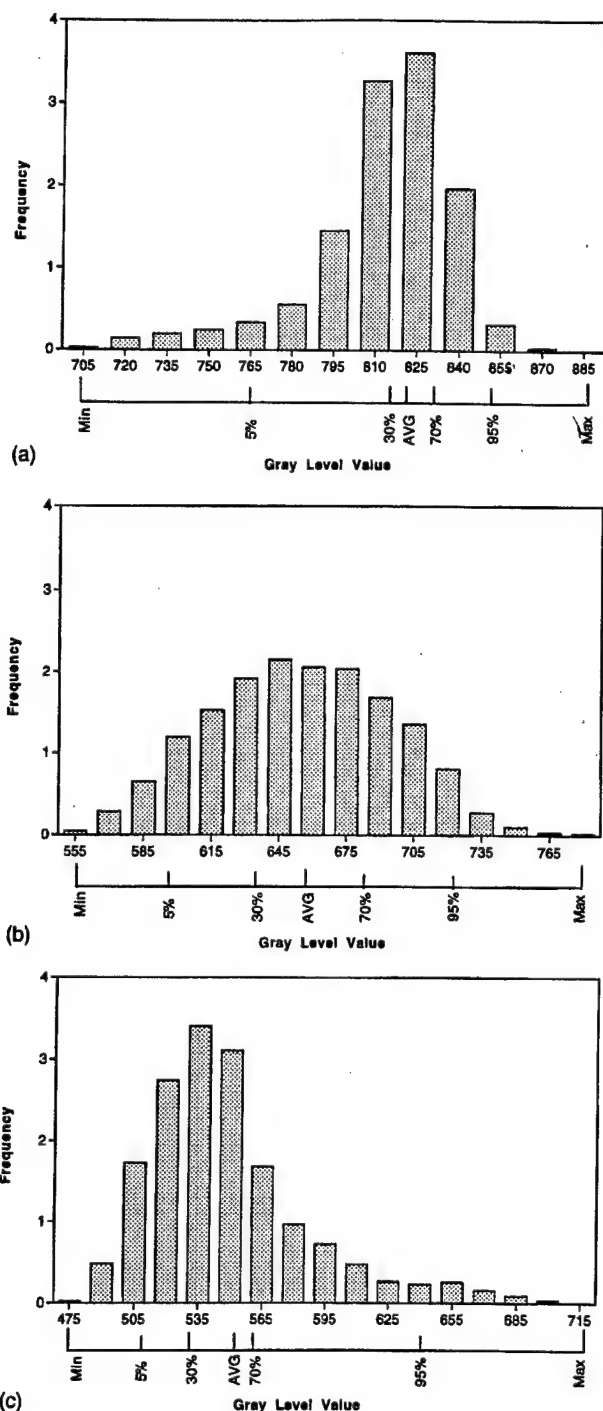*b. Features based on gray-level histogram analysis.* A dense ROI tends to have more pixels with high gray-level

(a)

(b)

(c)

FIG. 3. Gray-level histograms generated from (a) a dense ROI, (b) a mixed ROI, and (c) a fatty ROI.

TABLE I. List of the feature values obtained from the histograms (shown in Fig. 3) for a dense, a mixed, and a fatty ROI.

Features	The average values of the features		
	A dense ROI	A mixed ROI	A fatty ROI
Features based on the absolute value of the gray value			
MAX (gray level)	887	797	718
MIN (gray level)	705	555	473
AVG (gray level)	820	662	554
5% threshold (gray level)	765	597	507
95% threshold (gray level)	850	725	643
30% threshold (gray level)	814	639	533
70% threshold (gray level)	833	685	562
Features based on gray-level histogram analysis			
Balance1	0.55	0.97	1.70
Balance2	2.17	1.00	0.38
Skewness	−1.39	0.03	1.28

glandular tissue versus fat). As shown in Table I, a dense ROI should yield a negative value of skewness, a value less than one for balance1 and a value greater than one for balance2, whereas a fatty ROI should yield a positive value of skewness, a value greater than one for balance1 and a value less than one for balance2. A mixed ROI (half fatty and half dense) should yield a value close to zero for skewness, a value close to one for balance1, and balance2. The skewness measure has been studied by Byng *et al.*[23,33] to evaluate percent mammographic density in the breast. The balance1 measure has been studied by Tahoces *et al.*[30] to classify mammographic patterns into Wolfe patterns. We investigated two balance measures (i.e., balance1 and balance2) at different thresholds of the gray-level histogram to quantify the balance of the histogram.

(8) *Balance1*: (95% threshold-AVG)/(AVG-5% threshold).[30]

(9) *Balance2*: (70% threshold-AVG)/(AVG-30% threshold).

(10) *Skewness*: $m_3/m_2^{3/2}$, where

$$m_k = \sum_{i=0}^{G_h} n_i(i-\bar{i})^k/N,$$

$$N = \sum_{i=0}^{G_h} n_i, \quad \bar{i} = \sum_{i=0}^{G_h} n_i i,$$

and n_i is the number of occurrences of gray-level value i. G_h is the highest gray-level value in the ROI.[33]

c. Features based on spatial relationship among gray levels. Two features (coarseness and contrast) based on the spatial relationship among gray levels were investigated to characterize the texture patterns in the ROI. Coarseness and contrast were first proposed by Amadasun *et al.*[35] and have been used to characterize Wolfe patterns by Tahoces *et al.*[30] The mathematical definitions of the two texture features are given below. The coarseness of a texture is defined by the amount of local variation in gray level. The contrast of a

values (low optical density), yielding a gray-level histogram skewed to the left, as shown in Fig. 3(a). A fatty ROI tends to have more pixels with low gray-level values (high optical density), yielding a gray-level histogram skewed to the right, as shown in Fig. 3(c). Features such as skewness and balance (defined below) of a histogram relative to the mean can be used to quantify the ratio of pixels with high gray-level values to those with low gray-level values relative to the mean, thereby approximating the local tissue composition (fibro-

texture is defined by the amount of differences among all gray levels in the ROI and the amount of local variation in the gray level presented in the ROI. Notice that the contrast measure is determined by two terms: the gray-level differences in a ROI weighted by the amount of local variation. Thus, ROIs that have similar gray-level differences may have different contrast depending on the local variation in the ROIs. Conversely, ROIs that have the same amount of local variation may have different contrast depending on the gray-level differences in the ROIs.

(11) *Coarseness*: local uniformity,[35]

$$\text{COS} = \left[\sum_i^{G_h} p_i s(i) \right]^{-1}.$$

(12) *Contrast*: local contrast,[35]

$$\text{CON} = \left[\frac{1}{N_g(N_g-1)} \sum_{i=0}^{G_h} \sum_{j=0}^{G_h} p_i p_i (i-j)^2 \right] \left[\frac{1}{n^2} \sum_{i=0}^{G_h} s(i) \right],$$

where N_g is the total number of different gray levels present in the ROI, G_h is the highest gray-level value in the ROI, p_i is the probability of occurrence of gray-level value i, N is the width of the ROI, d is the neighborhood size (half of the operating kernel size), $n = N - 2d$, and the ith entry of s is given by

$$S(i) = \begin{cases} \sum |i - A_i|, & \text{for } i \in \{N_i\}, \quad \text{if } N_i \neq 0, \\ 0, & \text{otherwise,} \end{cases}$$

in which $\{N_i\}$ is the set of pixels having gray level i,

$$A_i = \frac{1}{W-1} \sum_{p=-d}^{d} \sum_{p=-d}^{d} f(x+p, y+q)$$

$(p,q) \neq (0,0)$ to exclude(x,y),

$W = (2d+1)^2 \quad (d=1)$.

d. Features based on Fourier transform analysis. The texture properties in each ROI were also analyzed from the two-dimensional Fourier transform. Background-trend correction was performed within the ROI prior to the application of the Fourier transform in order to reduce the contribution of variation from the gross anatomy of the breast background (low-frequency component).[36] The root-mean-square (RMS) variation and first moment of power spectrum (FMP) from the Fourier transform, as defined below,[37] were calculated to quantify the magnitude and spatial frequency content of the fine underlying texture in the ROI after the background trend correction. The RMS variation and the first moment of power spectrum have been investigated by Katsuragawa *et al.*[36] to analyze interstitial disease in chest radiographs, by Tahoces *et al.*[30] to classify Wolfe patterns in mammograms and by Caligiuri *et al.*[38] to characterize bone textures in lateral spine radiographs.

(13) *RMS variation*: root mean square of power spectrum,

$$\text{RMS} = \sqrt{\int \int |F(u,v)|^2 \, du \, dv}.$$

(14) *FMP*: first moment of power spectrum,

$$\text{FMP} = \int \int \sqrt{u^2 + v^2} |F(u,v)|^2 \, du \, dv \Big/ \int \int |F(u,v)|^2 \, du \, dv,$$

where $F(u,v)$ is the Fourier transform of the background corrected ROI.

2. Selection of computer-extracted mammographic features

a. Classification of BRCA1/BRCA2-mutation carriers and cases at low risk. We examined the computer-extracted features of the 15 BRCA1/BRCA2-mutation carriers and 143 ''low-risk'' women as one approach for relating mammographic features to breast cancer risk. In this approach, the ability of each individual computer-extracted feature was first evaluated using receiver operating characteristic (ROC) methodology[39,40] in the task of distinguishing between BRCA1/BRCA2-mutation carriers and the low-risk women. In the ROC analysis, the individual features were used as the decision variables. The area under the ROC curve (A_z) was used as an index to indicate the ability of the individual features in distinguishing between the 15 BRCA1/BRCA2-mutation carriers and the 143 ''low-risk'' women.

Next, stepwise linear discriminant analysis[41] was employed to select useful features from the 14 computer-extracted features. The stepwise linear discriminant analysis was accomplished in two steps. First, a stepwise feature selection was performed to identify useful features. Second, the selected features were used to determine the coefficient of each feature variable in the discriminant function to achieve maximum separation between the two groups. The discriminant function is formulated by a linear combination of the feature variables (the computer-extracted features). The criterion used to choose the best features in the stepwise procedure is to minimize the ratio of the within-group sum of squares to the total sum of the squares of the distribution of discriminant scores (Wilks' lambda). A detailed discussion of the underlying statistical theory for the stepwise procedure using the Wilks' lambda criterion is given in the literature.[41] The ability of the linear discriminant function, which merged the selected features, in distinguishing between the mutation carriers and the ''low-risk'' women, was also evaluated using ROC analysis. The discriminant score of each case from the linear discriminant function was used as the decision variable in the ROC analysis.

b. Correlation of mammographic features with risks as estimated from the Gail and the Claus models. In order to relate mammographic features to breast cancer risk, we employed linear regression analysis[42] to merge computer-extracted features along with age into a regression function to predict risk, as estimated from either the *Gail* model or the

TABLE II. Performance of 14 computer-extracted features in differentiating between the 15 BRCA1/BRCA2-mutation carriers and the ''low-risk'' cases in the entire database and the age-matched group in terms of A_z.

Features	Avg. value (mutation)	Avg. value (low risk)	A_z (entire group)	A_z (age matched)
Features based on the absolute value of the gray value				
MAX (gray level)	838	783	0.68 ± 0.06	0.69 ± 0.08
MIN (gray level)	561	517	0.59 ± 0.08	0.53 ± 0.09
AVG (gray level)	729	641	0.76 ± 0.06	0.71 ± 0.08
5% threshold (gray level)	578	570	0.74 ± 0.06	0.69 ± 0.08
95% threshold (gray level)	794	771	0.75 ± 0.06	0.71 ± 0.08
30% threshold (gray level)	624	618	0.76 ± 0.06	0.73 ± 0.07
70% threshold (gray level)	755	663	0.75 ± 0.06	0.72 ± 0.08
Features based on gray-level histogram analysis				
Balance1	0.90	1.09	0.70 ± 0.05	0.73 ± 0.07
Balance2	1.13	0.84	0.75 ± 0.05	0.80 ± 0.06
Skewness	−0.46	0.13	0.82 ± 0.04	0.87 ± 0.05
Features based on spatial relationship among gray levels				
Coarseness	0.000 65	0.000 48	0.72 ± 0.06	0.73 ± 0.07
Contrast	0.000 33	0.000 43	0.73 ± 0.06	0.74 ± 0.07
Features based on Fourier analysis				
FMP (cycles/mm)	7.09	7.10	0.74 ± 0.07	0.69 ± 0.08
Rms variation	24.81	20.21	0.70 ± 0.07	0.63 ± 0.08
Average A_z of the 14 features			0.73 ± 0.05	0.72 ± 0.08

Claus model. Both the lifetime risk (the risk of developing breast cancer up to age 70) and the 10 year risk (the risk of developing breast cancer in the next 10 years), as estimated from the models, were used as risk indices in the regression analysis.

The objective of regression analysis is to develop an equation that ''fits'' to observed variables, i.e., the risks estimated from the Gail or the Claus models. Stepwise regression was undertaken to identify from the 14 features, along with age, the most useful features to be used as the predictors in the regression function. The selected features were used to determine the regression coefficients in the regression function to achieve the minimum square difference between the observed variables and the estimated risk from the regression function. The forward stepwise procedure in MINITAB[43] was employed to select features. The criterion used in the feature selection is based on a measure (F^*-statistic) of the reduction in the variation of the observations around the fitted regression line. A detailed discussion of the underlying statistical theory can be found in the literature.[43] Four different regression functions were obtained for the four different observed variables, 10 year risk, and lifetime risk, as estimated from the *Gail* and the *Claus* models.

III. RESULTS

A. Mutation carriers and the low-risk women

Table II lists the A_z values indicating individual performance levels of the 14 features in the task of distinguishing between the BRCA1/BRCA2-mutation carriers and the low-risk cases in the entire group and the age-matched group. As shown in Table II, the majority of the features yield an A_z value greater than 0.70 in distinguishing between the mutation carriers and the ''low-risk'' cases in both the entire

group and the age-matched group. No consistent increases or decreases in the A_z values of the 14 individual features were observed when these features were applied to the age-matched group. The average of the A_z values from the 14 features obtained from the age-matched group ($A_z=0.72$) is similar to that obtained from the entire group ($A_z=0.73$). This suggests that the slight difference in age distribution between the BRCA1/BRCA2-mutation carriers and the ''low-risk'' cases does not have a strong influence on the performance of these individual features for this database.

The average values of individual features were calculated for the mutation cases and the ''low-risk'' cases only (Table II). The average value of the features based on gray level indicate that the selected ROIs corresponding to the mutation carriers yield higher gray-level values than those of the ''low-risk'' cases; the average value of the skewness and balance features show that the selected ROIs corresponding to the mutation carriers tend to have more pixels with high gray-level values relative to the pixels with low gray-level values than those corresponding to the ''low-risk'' cases. The average value of the texture features indicate that mammographic patterns of the mutation carriers tend to be coarser in texture and lower in contrast than do those of the ''low-risk'' cases. Figure 4 shows the distribution of the mutation carriers and the ''low-risk'' cases in terms of selected features: (a) RMS variation versus FMP and (b) coarseness versus skewness.

Four features were selected from the stepwise feature selection procedure for the classification of the mutation carriers and the ''low-risk'' cases. They are skewness, coarseness, contrast, and balance2. The linear discriminant function yielded an A_z of 0.91 in classifying the 15 BRCA1/BRCA2-mutation carriers and the 143 ''low-risk'' women.
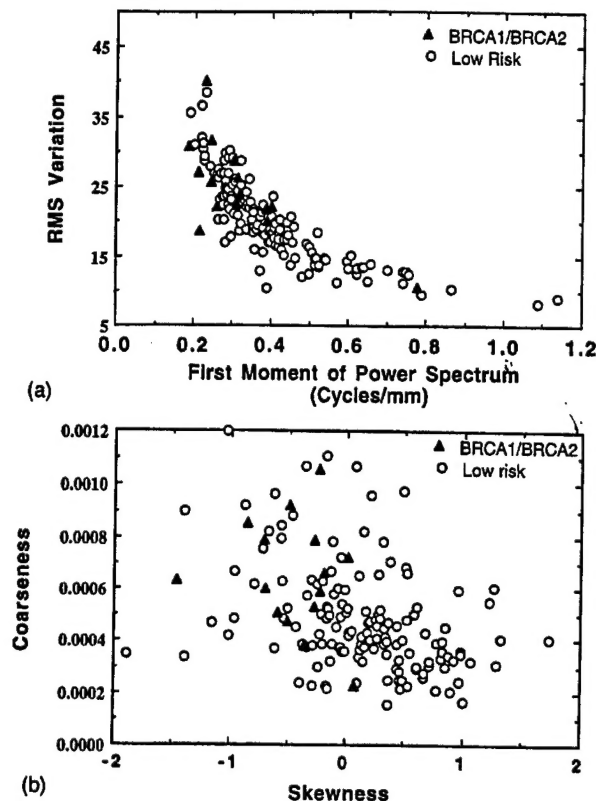
FIG. 4. Scattering plots of the BRCA1/BRCA2-mutation carriers and low-risk cases in terms of (a) RMS variation and FMP and (b) coarseness and skewness.

B. Correlation with the *Gail* and *Claus* models

Since the *Claus* model was designed to assess risk for women who have a family history of breast cancer, only 143 cases (dataset A) out of 341 cases (excluding BRCA1/BRCA2-mutation carriers) had such complete information, as required by the *Claus* model. Three hundred and three cases (dataset B) had complete information, as required by the *Gail* model. Datasets A and B were used to establish

models in predicting the lifetime risk and ten-year risk indices, as estimated from the *Claus* model and the *Gail* model, respectively. Dataset A and dataset B are overlapping subsets from the entire database. Thus, the cases used in establishing the linear regression functions to predict risk as estimated from the *Gail* model and the *Claus* model were different.

Since it is an important risk factor, age was used along with the mammographic features in the feature selection procedure. The stepwise feature selection procedure was performed on each of the two datasets and the corresponding models. A total of four sets of features were selected for the two risk indices (i.e., the lifetime risk or the ten-year risk) as estimated from the two models. The selected features along with their correlation coefficients are listed in Table III. The correlation coefficient (r) was calculated to evaluate the ability of the regression function using the selected features in predicting risk as determined from the clinical models.

We observed the following phenomena from the linear regression functions listed in Table III. With two different risk indices (i.e., lifetime risk and ten-year risk) and different subsets of the database, similar mammographic features (with one feature different) were identified as important features to predict risk, as estimated from the two clinical models. The association between the risk and a given feature, as indicated by its corresponding correlation sign (the negative/positive signs) in the regression functions, is the same for each of the four computer-extracted features in the different functions. The association between individual mammographic features and risk as estimated from the *Gail* or *Claus* model indicates that women with dense breasts (the negative sign for skewness), coarse (positive sign for coarseness) and low contrast (negative sign for contrast) mammographic patterns tend to have a high risk of developing breast cancer. It should be noted that age was used in both the *Gail* and the *Claus* models to predict risk. Results from Table III show that the ten-year risk increases as age increases, while the lifetime risk decreases as age increases.

TABLE III. The linear regression models generated for the lifetime risk and ten-year risk as estimated from (a) the Claus model using 143 cases and (b) the Gail model using 303 cases. Note: skew, cos, rms, and con correspond to the features of skewness, coarseness, rms variation, and contrast, respectively.

(a) Correlation with the Claus model Dataset A (143 cases)			
Responses	Models	r	p value
Lifetime risk	$0.32 - 0.032$ skew $+ 0.003$ rms $- 261.83$ con $- 0.003$ age	0.55	$<0.000\,01$
Ten year risk	$-0.09 - 0.013$ skew $+ 0.002$ rms $- 100.52$ con $+ 0.004$ age	0.57	$<0.000\,01$

(b) Correlation with the Gail model Data set B (303 cases)			
Responses	Features	r	p value
Lifetime risk	$0.22 - 0.014$ skew $+ 77.10$ cos $- 97.4\,1$ con $- 0.002$ age	0.41	$<0.000\,01$
Ten-year risk	$-0.03 - 0.004$ skew $+ 34.51$ cos $- 38.31$ con $+ 0.002$ age	0.41	$<0.000\,01$

IV. DISCUSSIONS

We investigated two different methods, i.e., a classification method and a correlation method, to identify useful mammographic features that are associated with predictors of breast cancer risk. The selected mammographic features were based on the analysis of the ROI selected from one of the four routine mammographic views (MLO and CC views of left and right breasts) obtained for each patient, namely, the left CC view. We have studied whether the mammographic characteristics as described by the computer-extracted features from a single image are representative and sufficient for the estimation of breast cancer risk. In the study, the correlation of each individual feature extracted from the two projections (CC and MLO views) of the same breast (left) and the correlation of each individual feature extracted from the same projection (CC view) of the left breast and the right breast were evaluated. In our database of 356 cases, the correlation coefficients of the 14 features ranged from 0.66 to 0.85 between images from CC and MLO views of the left breast and from 0.61 to 0.78 between images from CC views of the left and right breasts. Byng *et al.* have studied the left–right symmetry and projection (MLO vs CC view) symmetry of two computer-extracted texture measures (skewness and fractal dimension).[44] In a database of 30 cases, they found that the correlation coefficients for the two measures ranged from 0.86 to 0.93. Results from their study and ours indicate that a representative characterization of mammographic texture patterns can be obtained from analyses of a single projection of one of the breasts.

We realize that the size of ROI used in our study is a limitation since the ROI represents different percentages of the breast area for women with different breast sizes. Incorporation of the breast size in the analysis is important. In the future, we plan to vary the size of the ROI used for different sizes of breasts. We did investigate the use of five ROIs of the same size (256 pixels by 256 pixels) within the breast region: one at the center of the breast and one on each corner of the centered one. The centers of the four ROIs at the corners vary from breast to breast, depending on the size of the breast. The average of each individual computer-extracted feature over the five ROIs, however, performed similarly or poorer than that from the ROI behind the nipple. Use of the entire breast area as the ROI is ideal for evaluating the percent density of breast. Studies by others[30,31,33] have used multiple ROIs to include more breast area in their analyses. The results from our study may best assess the texture of dense regions (as opposed to percent dense), which usually occurs behind the nipple. A future investigation will address this issue.

Prediction of the breast cancer risk is a rather difficult task since it involves many factors. In the classification study, BRCA1/BRCA2 mutation is the only risk factor that was considered. The problem with this approach is that a few women in the "low-risk" group may actually have the BRCA1/BRCA2 gene mutation but are not aware of its presence. It is estimated that about 3 in 1000 women in the United States today have inherited susceptibility to breast cancer.[11] The likely prior probability that the women in our "low-risk" group would harbor BRCA1/BRCA2 mutations is low enough because they had no family history of breast cancer warranting genetic testing, and they were regarded as low risk without having to perform genetic testing. In the correlation study, differences in the results as shown in Table III when using the two models should not be unexpected since the two models were designed from two different populations and use different risk factors.[10,11] Further, the *Gail* and the *Claus* models are based on selected risk factors,[10] though the risk factors used in the models are considered to be the major factors and they were intended to be used to predict an individual's overall risk. Other studies indicated that increased mammographic density associates with increased breast cancer risk that could not be explained by other risk factors.[23] Thus, in our study, it is not unexpected that our computer-extracted features are not strongly correlated with the risks, as estimated from the models based on selected risk factors.

Although the risk in this study was not calculated from the true observations of breast cancer incidence for the studied population, results from our study agree well with the findings by others,[32] who related two computer-extracted texture features (skewness and fractal dimension) directly to "true" breast cancer risk (observed risk), and found that both measures were useful in characterizing mammographic density and in predicting risk. We found that the two approaches we employed are useful in identifying important mammographic features, which were consistently selected in both approaches. Results from both methods suggest that women at high risk, i.e., BRCA1/BRCA2-mutation carriers or non-mutation carriers, tend to have dense breasts and their mammographic patterns tend to be coarse and of low contrast. In fact, the two methods served as a validation method for each other in terms of feature selection. They can be used potentially in the future as means to estimate risk associated with breast cancer based on the analysis of mammograms and integrated with other clinical models. To our knowledge, it is the first time that computerized analyses are performed to analyze mammographic patterns of BRCA1/BRCA2-mutation carriers, and our results show that similar mammographic patterns may exist for the high-risk women in general and for women who are BRCA1/BRCA2-mutation carriers, in particular, based on computerized analyses.

V. CONCLUSION

Useful computer-extracted mammographic features were identified to be associated with breast cancer risk from two different approaches. Similar mammographic characteristics were found for high-risk women who are either mutation carriers or nonmutation carriers. The performance of the computer-extracted features suggest that women who are at high risk (mutation carriers or no-mutation carriers) tend to have dense breasts and their mammographic patterns tend to be coarse and low in contrast.

ACKNOWLEDGMENTS

[a] Corresponding author: Department of Radiology, MC2026, The University of Chicago, 5841 S. Maryland Ave, Chicago, Illinois 60637. Telephone: 773-702-6778; fax: 773-702-0371; electronic mail: z-huo@uchicago.edu

[1] American Cancer Society, *Cancer Facts and Figures—1998*, New York, 1998, p. 20.

[2] C. L. Carter, C. Allen, and D. E. Henson, "Relation of tumor size, lymph node status, and survival in 24,740 breast cancer cases," Cancer (N.Y.) 63, 181–187 (1989).

[3] M. G. Clay, G. Hishop, L. Kan, I. A. Olivotto, and L. J. W. Burhenne, "Screening mammography in British Columbia 1988–1993," Am. J. Surg. 167, 490–492 (1994).

[4] S. Shapiro, "Periodic breast cancer screening in seven foreign countries," Cancer (N.Y.) 69, 1919–1924 (1992).

[5] C. R. Smart, R. E. Hendrick, J. H. Rutledge, and R. A. Smith, "Benefit of mammography screening in women aged 40–49: current evidence from randomized controlled trials," Cancer (N.Y.) 75, 1619–1626 (1995).

[6] J. M. Elwood, B. Cox, and A. K. Richardson, "The effectiveness of breast cancer screening by mammography in younger women, Online J. Curr. Clin. Trials, 1993; Doc:32.

[7] G. Dodd, "American Cancer Society guidelines on screening for breast cancer: an overview," Cancer (N.Y.) 42, 177–180 (1992).

[8] L. Tabar, G. Fagerberg, and R. H. Chen, "Efficacy of breast screening by age: new results from the Swedish two county trial," Cancer (N.Y.) 75, 1412–1419 (1995).

[9] R. A. Smith, "Screening women aged 40–49: where are we today?," JNCI 87, 1198–1199 (1995).

[10] M. H. Gail and J. Benichou, "Assessing the risk of breast cancer in individuals," in *Cancer Prevention*, edited by V. T. DeVita, S. Hellman, and S. A. Rosenberg (J.B. Lippincott, Philadelphia, 1992), pp. 1–15.

[11] E. B. Claus, N. Risch, and W. D. Thompson, "Autosomal dominant inheritance of early-onset breast cancer: Implications for risk prediction," Cancer (N.Y.) 73, 643–651 (1993).

[12] K. F. Hoskins, J. E. Stopfer, and K. A. Calzone, "Assessment and counseling for women with a family history of breast cancer," J. Am. Med. Assoc. 273, 577–586 (1995).

[13] B. B. Biesecker, M. Boehnke, K. Calzone, D. S. Markel, J. E. Garber, F. S. Collins, and B. L. Weber, "Genetic counseling for families with inherited susceptibility to breast and ovarian cancer," J. Am. Med. Assoc. 269, 1970–1974 (1993).

[14] J. M. Hall, M. K. Lee, and J. Morrow, "Linkage of early-onset familial breast cancer to chromosome 17q21," Science 250, 1684–1689 (1990).

[15] D. Malkin, F. P. Li, and L. C. Strong, "Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms," Science 250, 1233–1238 (1990).

[16] M-C. King, "Breast cancer genes: how many, where, and who are they?," Nature Genet. 2, 250–290 (1992).

[17] K. Offit, *Clinical Cancer Genetics: Risk Counseling and Management* (Wiley, New York, 1998).

[18] B. Newman, M. A. Austin, M. Lee, and M-C. King, "Inheritance of human breast cancer: Evidence for autosomal dominant transmission in high risk families," Proc. Natl. Acad. Sci. USA 85, 3044–3048 (1988).

[19] E. B. Claus, N. Risch, and W. D. Thompson, "Genetic analysis of breast cancer in the cancer and steroid hormone study," Am. J. Hum. Genet. 48, 232–242 (1991).

[20] J. P. Struewing *et al.* "The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among ashkenazi jews," N. Engl. J. Med. 336, 1401–1408 (1997).

[21] D. F. Easton, D. Ford, and T. D. Bishop, "Breast and ovarian cancer incidence in BRCA1-mutation carriers," Am. J. Hum. Genet. 56, 256–271 (1995).

[22] J. Wolfe, "Breast patterns as an index of risk for developing breast cancer," Am. J. Roentgenol. 126, 1130–1139 (1976).

[23] J. W. Byng, N. F. Boyed, E. Fishell, R. A. Jong, and M. J. Yaffe, "The quantitative analysis of mammographic densities," Phys. Med. Biol. 39, 1629–1638 (1994).

[24] N. F. Boyd, B. O'Sullivan, J. E. Campbell, E. Fishell, I. Simor, and G. Cooke, "Mammographic signs as risk factors for breast cancer," Br. J. Cancer 45, 185–193 (1982).

[25] J. Brisson, A. S. Morrison, and N. Khalid, "Mammographic parenchymal features and breast cancer in the Breast Cancer Detection Demonstration Project," J. Natl. Cancer Inst. 80, 1534–1540 (1980).

[26] E. Warner, G. Lockwood, M. Math, D. Tritchler, and N. F. Boyd, "The risk of breast cancer associated with mammographic parenchymal patterns: a meta-analysis of the published literature to examine the effect of method of classification," Cancer Detection Prevention 16, 67–72 (1992).

[27] N. F. Boyd, B. O. O'Sullivan, E. Fishell, I. Simor, and G. Cooke, "Mammographic patterns and breast cancer risk: methodologic standards and contradictory results," J. Natl. Cancer Inst. 72, 1253–1259 (1984).

[28] I. E. Magnin, F. Cluzeau, and C. L. Odet, "Mammographic texture analysis: an evaluation of risk for developing breast cancer," Opt. Eng. 25, 780–784 (1986).

[29] C. B. Caldwell, S. J. Stapleton, D. W. Holdsworth, R. A. Jong, W. J. Weiser, C. Cooke, and M. J. Yaffe, "Characterization of mammographic parenchymal pattern by fractal dimension," Phys. Med. Biol. 35, 235–247 (1990).

[30] P. Tahoces, J. Correa, M. Souto, L. Gomes, and J. Vidal, "Computer-assisted diagnosis: the classification of mammographic breast parenchymal patterns," Phys. Med. Biol. 40, 103–117 (1995).

[31] P. Taylor, S. Hajnal, M-H. Dilhuydy, and B. Barreau, "Measuring image texture to separate "difficult" from "easy" mammograms," Br. J. Radiol. 67, 456–463 (1994).

[32] J. W. Byng, M. J. Yaffe, G. A. Lockwood, L. E. Little, D. L. Tritchler, and N. F. Boyd, "Automated analysis of mammographic densities and breast carcinoma risk," Cancer (N.Y.) 88, 66–74 (1997).

[33] J. W. Byng, N. F. Boyd, E. Fishell, R. Jong, and M. J. Yaffe, "Automated analysis of mammographic densities," Phys. Med. Biol. 1996, 909–923 (1996).

[34] Z. Huo, M. L. Giger, O. Olopade, and E. D. Wolverton, "Computer-aided diagnosis: Breast cancer risk assessment from mammographic parenchymal patterns in digitized mammograms," in *Proceedings of the 3rd International Workshop on Digital Mammography*, edited by K. Doi, M. L. Giger, R. M. Nishikawa, and R. A. Schmidt, Chicago, 1996, pp. 191–194.

[35] M. Amadasum and R. King, "Texture features corresponding to texture properties," IEEE Trans. Syst. Man Cybern. 19, 1264–1274 (1989).

[36] S. Katsuragawa, K. Doi, H. MacMahon, L. Monnier-Cholley, T. Ishida, and T. Kabayashi, "Classification of normal and abnormal lungs with interstitial disease by rule-based method and artificial neural networks," J. Dig. Imaging 10, 108–114 (1997).

[37] A. K. Jain, *Fundamentals of Digital Image Processing* (Prentice–Hall, Englewood Cliffs, NJ, 1986).

[38] P. Caligiuri, M. L. Giger, M. J. Favus, H. Jia, K. Doi, and L. B. Dixon, "Computerized radiographic analysis of osteoporosis. Preliminary evaluation," Radiology 186, 471–474 (1993).

[39] C. E. Metz, "ROC methodology in radiologic imaging," Invest. Radiol. 21, 720–733 (1986).

[40] C. E. Metz, "Some practical issues of experimental design and data analysis in radiological ROC studies," Invest. Radiol. 24, 234–245 (1989).

[41] P. L. Lachenbruch, *Discriminant Analysis* (Hafner Press, London, 1975).

[42] W. L. Hays, *Statistics* (Harcourt Brace College, Philadelphia, 1994).

[43] MINITAB Reference Manual, State College, PA, Minitab, Inc., 1995.

[44] J. W. Byng, N. F. Boyed, L. Little, G. Lockwood, E. Fishell, R. A. Jong, and M. J. Yaffe, "Symmetry of projection in the quantitative analysis of mammographic images," Eur. J. Cancer Prev. 41, 909–923 (1996).